DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Center for Biologics Evaluation and Research, HFM-215
Division of Biostatistics and Epidemiology
1401 Rockville Pike, Suite 200S
Rockville, MD 20852-1448
(301) 827-6849 (phone)
(301) 827-3529 (fax)

Memorandum

SUBMISSION: *Statistical Review and Evaluation of BLA #98-1296*

FROM: *Vance Berger, Ph.D., HFM-215*

THROUGH: *Peter A. Lachenbruch, Ph.D., Chief, Biostatistics Branch, HFM-215*

TO:

HFD-851 / Dr. Miller
HFM-99 / DCC
HFM-210 / Dr. Ellenberg
HFM-215 / Dr. Berger
HFM-215 / Chron
HFM-220 / Dr. Varrichio
HFM-500 / Dr. Siegel
HFM-561 / Dr. Rider
HFM-570 / Dr. Weiss
HFM-579 / Dr. Green
HFM-582 / Dr. Schwieterman
HFM-582 / Dr. Siegel
HFM-588 / Ms. Sickafuse
HFM-650 / Dr. Tavarez-Pagen

DATE: *May 18, 1999*

PRODUCT: *Enbrel (Etanercept), a Human Tumor Necrosis Factor (TNF) Receptor p75-Fc Fusion Protein Produced by Recombinant DNA Technology in a Chinese Hamster Ovary (CHO) Mammalian Cell Expression System*

INDICATION: *Juvenile Rheumatoid Arthritis (JRA) Polyarticular Course*

APPLICANT: *Immunex*

1. BACKGROUND:

Etanercept is a competitive inhibitor of the binding of TNF to its cell surface receptors and thereby regulates the biologic activity of TNF. Much of the joint pathology in RA is mediated by proinflammatory molecules that are linked in a network controlled by TNF. The mechanism of action of etanercept is thought to be its competitive inhibition of TNF binding to cell surface TNFR, preventing TNF-mediated cellular responses by rendering TNF biologically inactive. Etanercept may also modulate biologic responses controlled by additional downstream molecules (e.g., cytokines, adhesion molecules, or proteinases) that are induced or regulated by TNF. Etanercept decreases the levels of soluble adhesion molecules (e.g., E-selectin and intercellular adhesion molecule-1 or ICAM-1) in RA patients. Treatment of RA patients with etanercept also decreased serum levels of IL-6, which is thought to be produced by the cytokine cascade initiated by TNF. ENBREL is contraindicated in patients with or at risk of sepsis syndrome. ENBREL was recently approved for use in the adult RA population. The present submission is for JRA, polyarticular course.

2. KEY STUDIES:

There was one key study, Study #16.0016, "Safety, Population Pharmacokinetics, and Efficacy of TNFR:Fc in Children with Juvenile Rheumatoid Arthritis". This was a two-part, multicenter study in pediatric patients with polyarticular course, active JRA. In the first part of the study, all patients were treated with enbrel in an open-label fashion. The objective of Part 1 was to study the safety and population pharmacokinetics of enbrel. In the second part of the study, Part 2, responders from Part 1 were randomized to either placebo or to continue enbrel 0.4 mg/kg, for four months or until disease flare, whichever came first. To be considered a responder from Part 1, patients had to demonstrate a response at Day 90 as defined by the JRA Definition of Improvement (DOI), which includes a 30% improvement in at least three of the six following criteria, with 30% worsening in not more than one of the six following assessments:

1. Physician's global assessment;

2. Patient/parent global assessment;

3. Number of active joints (swelling not due to deformity or joints with LOM plus pain and/or tenderness);

4. Number of joints with LOM (modified by sponsor to include LOM plus pain and/or tenderness);

5. Functional assessment by the Childhood Health Assessment Questionnaire, or CHAQ (Singh 1994);

6. ESR (Giannini, 1997).

Responses were also assessed at baseline and Days 15, 30, 60, and 90, using the JRA Definition of Improvement (Giannini, 1997). Patients were permitted to remain on a stable dose of a single NSAID and/or corticosteroid at a dose of # 0.2 mg/kg or 10 mg maximum.

The objective of Part 2 was to study disease flare incidence and the median time to flare. The primary efficacy endpoint of the trial was the proportion of patients developing a disease flare in the two study arms. Disease flare (a significant worsening of disease activity compared to Day 90) was defined as ≥ 30% worsening in three of the six JRA Core Set Criteria and ≥ 30% improvement in not more than one of the six JRA Core Set Criteria (above), with a minimum of two active joints (swollen or LOM + P/T). If global assessments were used to establish flare, they had to have worsened by at least two units. The definition of flare was developed from a sensitivity analysis of several definitions of flare using data from the placebo-controlled trial of MTX in JRA (Giannini 1992), and was also accepted by the invesigators and the FDA as a definition with face validity.

Secondary endpoints for the randomized portion of the study included time to flare and responses to Enbrel, defined by the JRA DOI. Additional response assessments included articular severity score, pain score, duration of morning stiffness, and C-reactive protein (CRP). Trained joint assessors who were not involved in the patient's clinical care and who were blinded to study treatment in Part 2 of the study performed the joint assessments, whereas physician global assessments were performed by the principal investigators.

The randomization in Part 2 was stratified by active joint count, with one strata for "few" (0, 1, or 2) and one for "many" (at least three). There was also blocking, with a block size of two, and the randomization lists of the two strata within a center were mirror images of each other. For example, in Site #514, the first patient in the "few" strata was to receive placebo. This *implies* that the first patient in the "many" strata was to receive enbrel (there was not a separate randomization for each of the two strata within a center).

3. RESULTS:

This review is concerned primarily with efficacy from Part 2. In Part 2, 51 patients were randomized, 26 to placebo and 25 to enbrel. During the review of the efficacy data from Part 2, certain issues surfaced. Among these were the following:

1. . . The potential for unblinding of Part 2 treatment assignments;

2. Questionable decisions regarding which patients were randomized into Part 2 (Patients ——— did not meet the loss of motion (LOM) criteria with a minimum of three joints with LOM and P/T at screen: All were enrolled in Part 1, and Patients ——— were also randomized during Part 2; Patients ——— all received too much corticosteroid in the study and should not have been enrolled, but among these four patients, only Patient—— was not enrolled onto Part 2);

3. The randomization of certain patients (Patients ——————) from the wrong strata (sometimes influencing treatments assigned), per Table 1 below (the letter in parentheses indicates the treatment to be assigned from the stratum in question, and the number in parentheses indicates the probability one would ascribe to that treatment being assigned if they were aware of the previous set of treatment assignments):

TABLE 3.1: PATIENTS RANDOMIZED FROM THE WRONG STRATA

PATIENT #	PROPER STRATUM	ACTUAL STRATUM
	few (P, 1.00)	many (P, 1.00)
	many (E, 1.00)	few (P, 1.00)
	few (E, 1.00)	many (E, 0.50)
	few (P, 1.00)	many (E, 0.50);

4. Variable wash-out periods (Patient —— washed off methotrexate for only 13 days, instead of the 14 days specified by the protocol, resulting in Patient —— being randomized before Patient —— Patient —— washed off methotrexate for only 13 days, instead of the 14 days specified by the protocol, resulting in Patien —— being randomized before Patient ——, likewise, Patient —— should have been randomized before Patient ——, and would have been had Patient —— had proper wash-out; Patients —————— should have been randomized on the same date, but Patient —— was randomized a day earlier);

5. The failure to record the order in which patients were randomized, resulting in a need to use the time stamps on faxes to make this determination (but not all time stamps were readable, for example for Patient —— who was randomized on the same day, 3/11/98, as Patient —— and some faxes have multiple time stamps, resulting in imperfect information regarding the order in which patients were randomized);

6. Two reversals in the order of randomization (Patients —————— were both randomized on 12/23/97, the fax for Patien —— came at 2:08 PM, and the fax for Patient —— came at 3:33 PM, yet Patient —— was randomized first; Patients —————— were both randomized on 11/25/97, the fax for Patient —— came at 3:24 PM, and the fax for Patient —— came at 3:25 PM, yet Patient —— was randomized first);

7. Statistically significant imbalances at baseline between treatment groups (patients in the enbrel arm were younger, mean 8.9 years vs. 12.2 years, p=0.0026; less likely to be Caucasian, 56% vs. 88%, p=0.022; and had a lower mean weight, 34 kg vs. 43 kg, p=0.027).

The demographic imbalances between arms, coupled with errors in stratification, were the rationale for examining the possibility of selection bias in Part 2. With a block size of two, any unblinding of the first treatment allocation in a block would necessarily reveal the next treatment to be assigned, and this would also allow for prediction of treatment assignments to be made in the other stratum, because of the fact that the two stratum within a center had randomization from mirror-image lists. The predictability of treatment allocations would constitute a violation of allocation concealment,

and could lead to selection bias (Schultz, 1995). The result of selection bias would be that the groups were not balanced at randomization. If there was selection bias, then it would most likely have been at certain sites only, as selection bias is a "within-site phenomenon" when randomization is performed within site, as it was for this study. Consequently, our search for selection bias must necessarily consist of a series of searches, one per center. This effort is then somewhat hampered by the small numbers of patients enrolled at most centers, which made it difficult to formulate a comprehensive plan to detect selection bias.

The approach taken was as follows. The set of intended randomization codes, for each site, was used to determine the treatment assignment that would have been made, for each patient (even those patients who were not randomized), had the randomization been made according to either strata (the correct strata or the incorrect strata). Viewing the outcome of the randomization of a given patient as trichotomous (randomized according to the correct strata, not randomized at all, or randomized according to the wrong strata), what we are looking for is any predictability of this outcome by any knowledge of the treatment to be assigned (or the probability that one would ascribe, with knowledge of the previous treatment allocations, to the probability of either treatment being assigned to the current patient). The randomization data appear in Table 2 below.

TABLE 3.2: RESULTS OF RANDOMIZATION, BY CENTER

Site	Number Misrandomized	Number Not Randomized	Number Properly Randomized*	Total
31	0	2	8	10
174	0	5	1	6
182	0	3	4	7
242	0	1	4	5
502	1	2	2	5
503	2	1	12	15
504	1	2	6	9
506	0	0	2	2
514	0	2	8	10
Total	4	18	47	69

*This number includes all patients randomized from the correct strata, whether or not they should have been randomized during Part 2 of the study.

Site #174 is the only site to not randomize more patients then they did randomize. In fact, only one of six patients at Site #174 was randomized. This may reflect a differential interpretation of the criteria upon which randomization to Part 2 was based (i.e., response in Part 1).

Only six centers have adequate numbers in at least two categories to allow for any chance to detect selection bias if it were there. No effort can be made to detect selection bias at Sites #174 (only one patient enrolled), #242 (only one not enrolled), and #506 (all enrolled). Of the other six, three (Sites #502, #503, and #504) had mis-randomization of at least one patient to the wrong stratification

list. The other three are #31, #182, and #514. I find no evidence of selection bias at these last three sites (nor can it necessarily be ruled out).

At Site #502, Patient —— was misrandomized. This patient was an outlier, with a weight of 124.7 and a BSA of 2.41014, both extremely high. However, no change in treatment assigned, or even in the probability with which each treatment was to be assigned, resulted. This is because as the second patient to be randomized from the many stratum and the first patient to be randomized from the few stratum, the opposite treatment from the first treatment assigned from the many stratum had to be assigned. With knowledge that the first patient randomized to the many stratum, Patient —— was to receive enbrel, placebo was certain to be assigned to Patient —— regardless of the stratum from which the randomization was performed. However, it is curious that Page 12 of a fax from Dr. Mary Lange dated 1/29/99 indicates that the first patient randomized to the many strata at Site #502 should have received enbrel. In fact, Patient —— was the first patient randomized to the many stratum at Site #502 (12/3/97, 12:42), and received placebo. Patient —— was also randomized on 12/3/97 (12:44), to the many stratum, and received enbrel.

Had Patient —— been properly randomized, from the many stratum, then enbrel would have been certain to be administered. By switching strata, placebo was certain to be administered, and in fact was. In this case, the switch affected the treatment assignment.

Had Patient —— been properly randomized, from the few stratum, then enbrel would have been certain to be administered. By switching strata, placebo now had a 50% chance to be administered, but in fact enbrel turned out to be administered. In this case, the switch did not affect the treatment assignment, but did affect the probability with which enbrel was to be assigned.

Had Patient —— been properly randomized, from the few stratum, then placebo would have been certain to be administered. By switching strata, placebo now had only a 50% chance to be administered, and in fact enbrel was administered. In this case, the switch affected the treatment assignment, as well as the probability with which enbrel was to be assigned.

If there was selection bias, then a significant between-group difference in flare rate at the end of the study could not necessarily be attributed to enbrel, because it could just as well be explained by baseline imbalances in prognostic factors (Proschan, 1994). As such, the approach taken to evaluate robustness was to determine the number of patients required to switch treatment groups to break the observed statistically significant difference. The primary efficacy variable was flare rate. The data submitted by the sponsor gave a 2x2 table of flare rates as follows:

	Flare	No Flare	Total
Placebo	21	5	26
Enbrel	7	18	25.

However, further discussions with the FDA Medical Officer (Dr. Rider) revealed that one patient per arm (Patients [redacted] was classified incorrectly as a flare. There was agreement that the flare data should have been as follows:

	Flare	No Flare	Total
Placebo	20	6	26
Enbrel	6	19	25.

For this data set, Fisher's exact test yields p=0.0002 (one-sided). This means that enbrel, compared to placebo, tended to be significantly associated with lower flare rates. The question before us, in synthesizing this result with the potential for selection bias, is how many patients would be required to switch treatment groups, while maintaining their respective response rates, to break this observed statistical significance at one-sided level 0.025? We could attempt to determine which patients were randomized based on selection bias, and consequently should have been in the other treatment group. However, we will not pursue this approach for two distinct reasons. First, there is necessarily uncertainty regarding whether or not selection bias compromised any particular allocation. Second, even if we were able to determine those allocations that were compromised due to selection bias, there would still be a rippling effect, in that subsequent treatment allocations would also be reversed. For these reasons, we do not consider the analyses after switching certain patients across treatment groups, but rather consider the extreme case of switching those patients without flares on enbrel and those patients with flares on placebo. With one such switch, the flare data would be as follows:

	Flare	No Flare	Total
Placebo	19	7	26
Enbrel	7	18	25.

For this data set, Fisher's exact test yields p=0.0015 (one-sided). With two such switches, the flare data would be as follows:

	Flare	No Flare	Total
Placebo	18	8	26
Enbrel	8	17	25.

For this data set, Fisher's exact test yields p=0.0083 (one-sided). With three such switches, the flare data would be as follows:

	Flare	No Flare	Total
Placebo	17	9	26
Enbrel	9	16	25.

For this data set, Fisher's exact test yields p=0.0340 (one-sided), and we would no longer have statistical significance at one-sided level 0.025, with p=0.0340. Consequently, switching as few as three patients across treatment groups could break the observed statistical significance in flare rate.

We perform the same type of sensitivity analysis in the JRA DOI data across treatment groups. The Day 210 JRA DOI response rates, compared to Day 90 (the day of randomization), were as follows:

Reviewer's Table 4A1: JRA DOI Efficacy Data, Phase II Study 16.0004, Week 2					
p=0.0006	No JRA DOI30	JRA DOI 30	JRA DOI 50	JRA DOI 70	Total
Placebo	17	3	1	5	26
Enbrel	5	2	7	11	25
Total	22	5	8	16	51

The Smirnov test (Berger, Permutt, and Ivanova, 1998) yields:

	Asymptotic	Exact
One-Sided	p=0.0022	p=0.0006
Two-Sided	p=0.0045	p=0.0008.

We are particularly interested in the exact one-sided p-value of 0.0006. Again, the question before us is how many patients would be required to switch treatment groups, while maintaining their respective response rates, to break this observed statistical significance at one-sided level 0.025? Again, we consider the extreme case of switching those patients with the best responses (JRA DOI 70) to enbrel and those patients with the worst response to placebo (not even JRA DOI 30). With one such switch, the Day 210 JRA DOI data would be as follows:

	<30	30-50	50-70	>70	Total
Placebo	16	3	1	6	26
Enbrel	6	2	7	10	25.

The Smirnov test would then yield:

	Asymptotic	Exact
One-Sided	p=0.0136	p=0.0043
Two-Sided	p=0.0271	p=0.0062,

and we would still have statistical significance at one-sided level 0.025, with p=0.0043. With two such switches, the Day 210 JRA DOI data would be as follows:

	<30	30-50	50-70	>70	Total
Placebo	15	3	1	7	26
Enbrel	7	2	7	9	25.

The Smirnov test would then yield:

	Asymptotic	Exact
One-Sided	p=0.0599	p=0.0211
Two-Sided	p=0.1198	p=0.0329

and we would still have statistical significance at one-sided level 0.025, with p=0.0211. With three such switches, the Day 210 JRA DOI data would be as follows:

	<30	30-50	50-70	>70	Total
Placebo	14	3	1	8	26
Enbrel	8	2	7	8	25.

The Smirnov test would then yield:

	Asymptotic	Exact
One-Sided	p=0.1935	p=0.0750
Two-Sided	p=0.3842	p=0.1252,

and we would no longer have statistical significance at one-sided level 0.025, with p=0.0750. Once again, we see that switching as few as three patients across treatment groups could break the observed statistical significance.

5. SUMMARY

Enbrel appears to be efficacious, based on both the flare rate data and the JRA DOI data. However, there were methodological problems with the study.

6. CONCLUSIONS

The data are suggestive of an efficacy claim for enbrel.

REFERENCES

Berger, V., Permutt, T., and Ivanova, A. (1998). "The Convex Hull Test for Ordered Categorical Data", *Biometrics* **54**, 4, 1541-1550.

Giannini, E. H., Brewer, E. J., Kuzmin, N., et al. (1992). "Methotrexate in Resistant Juvenile Rheumatoid Arthritis. Results of the USA-USSR Double-Blind, Placebo-Controlled Trial", *N. Engl. J. Med* **326**, 1043-1049.

Giannini, E. H., Ruperto, N., Ravelli, A., et al. (1997). "Preliminary Definition of Improvement in Juvenile Arthritis", *Arthritis Rheum* **40**, 1202-1209.

Proschan, M. (1994). "Influence of Selection Bias on Type I Error Rate Under Random Permuted Block Designs", *Statistica Sinica* **4**, 219-231.

Schulz, K. F. (1995). "Subverting Randomization in Controlled Trials", *JAMA* **274**, 1456-1458.

Singh, G., Athreya, B. H., Fries, J. F., and Goldsmith, D. P. (1994). "Measurement of Health Status in Children with Juvenile Rheumatoid Arthritis", *Arthritis Rheum* **37**, 1761-1769.